WHAT IS CLAIMED IS:

1. A method of detecting new events comprising the steps of:

determining at least one story characteristic based on at least one of: an average story similarity story characteristic and a same event-same source story characteristic;

determining a source-identified story corpus, each story associated with at least one event;

determining a source-identified new story associated with at least one event;

determining story-pairs based on the source-identified new-story and each story in the source-identified story corpus;

determining at least one inter-story similarity metric for the story-pairs;

determining at least one adjustment to the inter-story similarity metrics based on at least one story characteristic; and

determining if the event associated with the new story is similar to the events associated with the source-identified story corpus based on the inter-story similarity metrics and the adjustments.

2. The method of claim 1, wherein the inter-story similarity metric is adjusted based on at least one of subtraction and division.

3. The method of claim 1, wherein the inter-story similarity metric is at least one of a probability based inter-story similarity metric and a Euclidean based inter-story similarity metric.

4. The method of claim 3, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto, a KL divergence and a clarity distance based metric.

5. The method of claim 3, wherein the Euclidean based similarity metric is a cosine-distance based metric.

6. The method of claim 1, wherein the inter-story similarity metrics are determined based on a term frequency-inverse story frequency model.

7. The method of claim 1, wherein the inter-story similarity metrics are comprised of: at least one story frequency model; and at least one event frequency model combined using terms weights.

8. The method of claim 1, wherein the inter-story similarity metrics are

comprised of at least one story frequency model; and at least one story characteristic frequency model combined using terms weights.

9. The method of claim 8, where the adjustments based on the story characteristics are applied to the term weights.

10. The method of claim 8, where the adjustments based on the story characteristics are applied to the inter-story similarity metrics.

11. The method of claim 1, wherein the inter-story similarity metrics are comprised of at least one term frequency-inverse event frequency model and where the events are classified based on at least one of: story labels and a predictive model.

12. The method of claim 8, wherein an event frequency is determined based on term *t* and *ROI* category *rmax* from the formula:

$$ef_{r\max}(t) = \max_{r \in R}(ef(r,t)).$$

13. The method of claim 8, wherein an the inverse event frequency is determined based on term *t*, and events *e* and *rmax* in the set of *ROI* categories from the formula: $IEF(t) = \log\left[\frac{N_{e,r\max}}{ef_{r\max}(t)}\right]$.

14. The method of claim 8, wherein an inverse event frequency is determined based on term *t*, categories *e,r* and *rmax* in the set of *ROI* categories and *P(r)*, the probability of *ROI r* from the formula:

$$IEF'(t) = \sum_{r \in R} P(r) \log\left[\frac{N_{e,r}}{ef(r,t)}\right].$$

15. The method of claim 1 further comprising the step of determining a subset of stories from the source-identified story corpus and the source-identified new story based on at least one story characteristic.

16. A system for detecting new events comprising:

an input/output circuit for retrieving source-identified new story and a source-identified story corpus, each story associated with at least one event;

a memory;

a processor for determining stories from the source-identified story corpus; and wherein the processor determines story-pairs based on the source -identified new story and each corpus story;

a similarity determining circuit for determining inter-story similarity information for the story-pairs;

a story characteristic adjustment circuit for determining adjustments to the inter-story similarity information based on at least one story characteristic; and

a new event determining circuit for determining a new event based on the inter-story similarity information and the story characteristic adjustments, and wherein the at least one story characteristic is based on at least one of: an average story similarity story characteristic and a same event-same source story characteristic;.

17. The system of claim 16, wherein the inter-story similarity information is adjusted based on at least one of subtraction and division.

18. The system of claim 16, wherein the inter-story similarity information is at least one of a probability based inter-story similarity information and a Euclidean based inter-story similarity information.

19. The system of claim 17, wherein the probability based inter-story similarity information is at least one of a Hellinger, a Tanimoto, a KL divergence and clarity distance based information.

20. The system of claim 17, wherein the Euclidean based similarity information is cosine-distance based information.

21. The system of claim 16, wherein the inter-story similarity information is determined based on a term frequency-inverse story frequency model.

22. The system of claim 16, wherein the inter-story similarity information is comprised of: at least one story frequency model; and at least one event frequency model combined using terms weights.

23. The system of claim 16, wherein the inter-story similarity information is comprised of at least one story frequency model; and at least one story characteristic frequency model combined using terms weights.

24. The system of claim 23, wherein the adjustments based on the story characteristics are applied to the term weights.

25. The system of claim 23, wherein the adjustments based on the story characteristics are applied to the inter-story similarity information.

26. The system of claim 16, wherein the inter-story similarity information

is comprised of at least one term frequency-inverse event frequency model and where the events are classified based on at least one of: story labels and a predictive model.

27. The system of claim 23, wherein an event frequency is determined based on term *t* and *ROI* category *rmax* from the formula:

$$ef_{r\max}(t) = \max_{r \in R}(ef(r,t)).$$

28. The system of claim 23, wherein an inverse event frequency is determined based on term *t*, and events *e* and *rmax* in the set of *ROI* categories from the formula: $IEF(t) = \log\left[\frac{N_{e,r\max}}{ef_{r\max}(t)}\right]$.

29. The system of claim 23, wherein an inverse event frequency is determined based on term *t*, categories *e,r* and *rmax* in the set of *ROI* categories and *P(r)*, and the probability of *ROI r* from the formula:

$$IEF'(t) = \sum_{r \in R} P(r) \log\left[\frac{N_{e,r}}{ef(r,t)}\right].$$

30. The system of claim 16 wherein the processor determines a subset of stories from the source-identified story corpus and the source-identified new story based on at least one story characteristic.

31. A carrier wave encoded to transmit a control program, useable to program a computer to detect new events, to a device for executing the program, the control program comprising:

instructions for determining at least one story characteristic based on at least one of: an average story similarity story characteristic and a same event-same source story characteristic;

instructions for determining a source-identified story corpus, each story associated with at least one event;

instructions for determining a source-identified new story associated with at least one event;

instructions for determining stories from the source-identified story corpus and the source-identified new story based on at least one story characteristic;

instructions for determining story-pairs based on the source-identified new-story and the set of stories based on the story characteristics;

instructions for determining at least one inter-story similarity metric for the story-pairs based on the source of the stories;

instructions for determining at least one adjustment to the inter-story similarity metrics based on at least one story characteristic; and

instructions for determining new events based on the inter-story similarity metrics and the adjustments.

32. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code usable to program a computer to detect new events comprising the steps of:

determining at least one story characteristic based on at least one of: an average story similarity story characteristic and a same event-same source story characteristic;

determining a source-identified story corpus, each story associated with at least one event;

determining a source-identified new story associated with at least one event;

determining stories from the source-identified story corpus and the source-identified new story based on at least one story characteristic;

determining story-pairs based on the source-identified new-story and the set of stories based on the story characteristics;

determining at least one inter-story similarity metric for the story-pairs based on the source of the stories;

determining at least one adjustment to the inter-story similarity metrics based at least one story characteristic; and

determining new events based on the inter-story similarity metrics and the adjustments.

33. A method of combining inter-story similarity information comprising the steps of:

determining P(sameROI(q,d)) based on the probability of story q and story d having the same ROI category;

determining $similarity_{IEF''}$, based on a similarity with no inverse event frequency influence; and

the formula:

$$similarity'(q,d) = P(sameROI(q,d)) * similarity_{IEF'}(q,d) + (1 - P(sameROI(q,d))) * similarity_{IEF''}(q,d)$$

34. The method of claim 33, wherein P(sameROI(q,d)) is based on the formula:

$$P(sameROI(q,d)) = \frac{N_{same}(similarity_{IEF''}(q,d))}{N_{same}(similarity_{IEF''}(q,d)) + N_{different}(similarity_{IEF''}(q,d))}$$

35. A method of detecting new events comprising the steps of:

determining a first source-identified story associated with at least one event;

determining a second source-identified associated with at least one event;

determining a story-pair based on the first source-identified story and the second source-identified story;

determining inter-story similarity between the first and second story based on at least one of: an event frequency model, story segmentation and a source-identified inter-story similarity metric.

36. The method of claim 35, wherein story segmentation is based on at least one of: topic, an adjacent window and an overlapping window.

37. A method of determining a predictive model for new event detection comprising the steps of:

determining a current story and corpus of stories each associated with at least one event;

determining cost information;

determining a multi-story similarity metric based on the current story and a plurality of at least two corpus stories;

determining new event decision model information;

determining new event information for the current story based on the new event decision model information and the multi-story similarity metric;

determining event training information;

determining new event decision model information based on the event training information, the cost information and a learner.